



# МАТЕМАТИЧЕСКАЯ ФИЗИКА, МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

УДК 519.6:004.8

## АНАЛИЗ МОДИФИЦИРОВАННОГО ГЕНЕТИЧЕСКОГО АЛГОРИТМА НА ОСНОВЕ ТЕОРИИ СХЕМ

## ANALYSIS OF MODIFIED GENETIC ALGORITHM BASED ON THE THEORY OF SCHEMES

**А.В. Глушак, В.А. Ломазов, Д.А. Петросов**  
**A.V. Glushak, V.A. Lomazov, D.A. Petrosov**

*Белгородский национальный исследовательский университет, Россия, 308015, г.Белгород, ул. Победы, 85*  
*Belgorod National Research University, 85 Pobedy St, Belgorod, 308015, Russia*

*Белгородский государственный аграрный университет им. В.Я. Горина, 308503, Белгородская обл.,  
Белгородский р-н, п.Майский, ул. Вавилова, 1*  
*Belgorod State Agricultural University by V.Y. Gorin, 1, Vavilova Street, p.Mayskiy, Belgorod district,  
Belgorod region, 308503, Russia*

*Белгородский государственный аграрный университет им. В.Я. Горина, 308503, Белгородская обл.,  
Белгородский р-н, п.Майский, ул. Вавилова, 1*  
*Belgorod State Agricultural University by V.Y. Gorin, 1, Vavilova Street, p.Mayskiy, Belgorod district,  
Belgorod region, 308503, Russia*

*E-mail: Glushak@bsu.edu.ru, vlomazov@yandex.ru, Scorpionss2002@mail.ru*

**Аннотация.** Приводится математическое обоснование модифицированного генетического алгоритма с применением турнирно-рулеточной селекции в виде аналога теоремы схем.

**Resume.** A mathematical study of the modified genetic algorithm with roulette-tournament selection in the form of analog scheme theorem is proposed.

**Ключевые слова:** генетический алгоритм, хромосома, популяция, селекция, генетические алгоритмы, теорема схем.

**Key words:** genetic algorithm, chromosome, population, selection, genetic algorithms, scheme theorem

## Введение

Генетические алгоритмы, представляющие собой биоинспирированную эвристическую процедуру направленного случайного поиска численного решения задачи дискретной оптимизации, получает в настоящее время всё большее распространение (см., например, [1 – 3]). При многоэкстремальности целевой функции, когда, например, метод конфигураций и другие традиционные методы прямого поиска нулевого порядка перестают быть эффективными и часто дают в качестве решения локальный экстремум, генетический алгоритм позволяет найти приближённое, удовлетворительное для многих практических задач решение, не прибегая к полному перебору элементов, хотя и конечного, но, как правило, большого по мощности пространства поиска.

В теории генетических алгоритмов используется словарь, заимствованный из естественной генетики. Мы будем придерживаться следующей терминологии (см., например, [1 – 3]):

каждое решение (особь) представимо битовой строкой (хромосомой) длины  $m$ :  
 $\alpha = (a_1, a_2, \dots, a_i, \dots, a_m), a_i \in \{0, 1\}$ ;

популяция  $P$  – набор хромосом, соответствующий пространству поиска:  $P = \{0, 1\}^m$ ;



генерация  $G(t)$  – часть популяции, представляющая собой новое поколение хромосом после каждого шага  $t$  ( $t = 0, 1, 2, \dots$ ) генетического алгоритма, размер генерации будем считать одинаковым и равным  $M$  для всех  $t$ :  $G(t) \subseteq P$ ,  $|G(t)| = M$ ;

родительский пул – часть текущей генерации  $G(t)$ , состоящая из хромосом, из которых путем скрещивания и мутации (генетические операторы) формируется новая генерация  $G(t+1)$ ;

приспособленность особи – целевая функция  $\mu(a) : P \rightarrow R$  задачи оптимизации, заданная на множестве хромосом.

Анализ генетических алгоритмов базируются на понятии схемы (шаблона), отражающем сходство между хромосомами. Схема  $s$  представляет собой строку длины  $m$  в алфавите  $P = \{0, 1, *\}$ , где  $*$  – символ безразличия. Хромосома  $a$  называется соответствующей схеме  $s = (s_1, s_2, \dots, s_i, \dots, s_m)$ , если  $a_i = s_i$  при  $s_i \neq *$  для всех  $i = 1, 2, \dots, m$ .

Отметим два важных понятия для схемы: порядок и определяющая длина схемы, которые используются при выводе фундаментальной теоремы генетических алгоритмов.

Порядок схемы  $o(s) = \sum_{i=1}^m \chi_i$ , где  $\chi_i = \begin{cases} 0, & s_i = *, \\ 1, & s_i \neq *. \end{cases}$  По-другому, порядок схемы – это её длина ми-

нус число символов безразличия. Определение порядка необходимо при вычислении вероятности выживания эталона при мутации.

Второе понятие – это определяющая длина схемы  $d(s) = \max\{i : \chi_i = 1\} - \min\{i : \chi_i = 1\}$ , которая представляет собой расстояние между первой и последней фиксированными позициями в строке. Это понятие дает оценку компактности информации, содержащейся в эталоне. Понятие определяющей длины важно при вычислении вероятности выживания эталона при скрещивании.

Итак, каждая хромосома  $a$  представляет потенциальное решение задачи дискретной оптимизации в конечном пространстве поиска. Эволюционный процесс представляет собой итерационную процедуру перехода от одной генерации хромосом к другой. При этом предполагается, что специальный способ перехода приводит к появлению в каждой последующей генерации более приспособленных особей чем в предыдущей генерации, т.е. к улучшению генерации, вплоть до нахождения близкого к оптимальному решения. Если дальнейшее выполнение итерационного процесса не приводит к улучшению уже достигнутого результата, то процесс прекращается.

В настоящее время при решении задач дискретной оптимизации во многих предметных областях широко используются различные модификации генетического алгоритма, в то время как математическое обоснование предположения об улучшении генерации (теорема схем) было известно (см. [4]) только для стандартного генетического алгоритма в случае рулеточной селекции. Для подтверждения возможности применения модификации генетического алгоритма в случае турнирно-рулеточной селекции при формировании родительского пула  $P$  (см. [5, 6]) использовались только методы имитационного моделирования, состоящие в проведении многочисленных вычислительных экспериментов.

В работе предложено математическое обоснование модифицированного генетического алгоритма с применением турнирно-рулеточной селекции в виде аналога теоремы схем.

### Основные результаты

При турнирно-рулеточной селекции случайным образом выбирается  $N$  хромосом ( $N$  – количество участников турнира) и наиболее приспособленная хромосома проходит в родительский пул с некоторой заданной вероятностью. Этот способ селекции предполагает попадание в родительский пул на первых этапах и не самых приспособленных хромосом, что способствует поиску оптимального решения во всём пространстве потенциальных решений.

**Случай**  $N = 2$ . Будем считать, что из двух участвующих в турнире хромосом  $a$  и  $b$  наиболее приспособленная хромосома, т.е. хромосома, для которой  $\mu(a) \geq \mu(b)$ , проходит в родительский пул

$P$  с вероятностью  $\frac{\mu(a)}{\mu(a) + \mu(b)}$ , а другая с вероятностью  $\frac{\mu(b)}{\mu(a) + \mu(b)}$ .



Хромосомы популяции, соответствующие схеме  $s$ , обозначим  $a_1, a_2, \dots, a_y$ ,  $y < M$ , а не соответствующие схеме  $s$  –  $b_1, b_2, \dots, b_{M-y}$ . Определим вероятность того, что хромосома, соответствующая схеме  $s$ , пройдёт в родительский пул  $P$ .

Если в турнир отобраны две хромосомы, соответствующие схеме  $s$ , то с вероятностью равной 1 в родительский пул пройдёт хромосома, соответствующая схеме  $s$ .

Если в турнир отобрана одна хромосома  $a_i$ ,  $1 \leq i \leq y$ , соответствующая схеме  $s$ , а другая  $b_j$ ,  $1 \leq j \leq M - y$ , не соответствующая схеме  $s$ , то искомая вероятность равна  $\frac{2y(M-y)}{M^2} \cdot \frac{\mu(a_i)}{\mu(a_i) + \mu(b_j)}$ , и средняя вероятность того, что хромосома  $a_i$ ,  $1 \leq i \leq y$ , соответствующая схеме  $s$ , пройдёт в родительский пул будет равной  $\frac{2}{M^2} \sum_{\substack{1 \leq i \leq y \\ 1 \leq j \leq M-y}} \frac{\mu(a_i)}{\mu(a_i) + \mu(b_j)}$ .

Количество хромосом родительского пула  $P$  соответствующее схеме  $s$  на шаге  $t$ , обозначим  $y(s, t)$ . Тогда математическое ожидание количества хромосом родительского пула на следующем шаге будет равно

$$E(y(s, t+1)) = y(s, t) \left( \frac{y^2(s, t)}{M^2} + \frac{2}{M^2} \sum_{\substack{1 \leq i \leq y(s, t) \\ 1 \leq j \leq M-y(s, t)}} \frac{\mu(a_i)}{\mu(a_i) + \mu(b_j)} \right) \quad (1)$$

Уравнение (1) назовём уравнением репродуктивного роста схемы  $s$  в рассматриваемом случае  $N = 2$  турнирно-рулеточной селекции.

Таким образом, при выполнении неравенства

$$\frac{y^2(s, t)}{M^2} + \frac{2}{M^2} \sum_{\substack{1 \leq i \leq y(s, t) \\ 1 \leq j \leq M-y(s, t)}} \frac{\mu(a_i)}{\mu(a_i) + \mu(b_j)} > 1$$

схема получает увеличенное число хромосом в следующую генерацию  $G$ .

В частности, из уравнения (1) следует неравенство

$$E(y(s, t+1)) \geq y(s, t) \left( \frac{y^2(s, t)}{M^2} + \frac{2y(s, t)(M - y(s, t))}{M^3} \cdot \frac{\bar{\mu}(s, t)}{\bar{F}(t)} \right), \quad (2)$$

где  $\bar{\mu}(s, t) = \frac{1}{y(s, t)} \sum_{i=1}^{y(s, t)} \mu(a_i)$  – средняя приспособленность хромосом, соответствующих схеме  $s$ , в момент  $t$ ,  $\bar{F}(t) = \frac{1}{M} \left( \sum_{i=1}^{y(s, t)} \mu(a_i) + \sum_{j=1}^{M-y(s, t)} \mu(b_j) \right)$  – средняя приспособленность хромосом всей популяции в момент  $t$ .

Как следует из неравенства (2), число строк в популяции, отвечающих схеме  $s$ , растёт, если

$$\frac{\bar{\mu}(s, t)}{\bar{F}(t)} > \frac{M(M + y(s, t))}{2y(s, t)}. \quad (3)$$

Следовательно, выполнение неравенства (3) является простым достаточным условием роста числа хромосом в следующую генерацию  $G$ .

Оценка математического ожидания  $E(y(s, t+1))$ , получаемая из уравнения (1) репродуктивного роста схемы  $S$  с учётом генетических операторов скрещивания и мутации для турнирно-рулеточной селекции выводится как и в классическом случае (см. [1 – 4]) и имеет вид



$$E(y(s, t+1)) \geq y(s, t) \left( \frac{y^2(s, t)}{M^2} + \frac{2}{M^2} \sum_{\substack{1 \leq i \leq y(s, t) \\ 1 \leq j \leq M-y(s, t)}} \frac{\mu(a_i)}{\mu(a_i) + \mu(b_j)} \right) \cdot \left( 1 - p_c \frac{d(s)}{m-1} \right) \cdot (1 - p_m)^{o(s)}, \quad (4)$$

где  $o(s)$  и  $d(s)$  соответственно порядок и определяющая длина схемы  $s$ ,  $p_c, p_m$  – вероятности скрещивания и мутации хромосом из родительского пула  $P$ .

Неравенство (4) определяет ожидаемое число сходных со схемой хромосом в последующих генерациях как функцию числа таких хромосом в текущей популяции  $y(s, t)$ , относительной пригодности схемы, её определяющей длины  $d(s)$  и порядка  $o(s)$ .

**Случай**  $N=3$ . Будем считать, что три участвующие в турнире хромосомы  $a, b$  и  $c$ , для которых  $\mu(a) \geq \mu(b) \geq \mu(c)$ , проходят в родительский пул  $P$  соответственно с вероятностями

$$\frac{\mu(a)}{\mu(a) + \mu(b) + \mu(c)}, \quad \frac{\mu(b)}{\mu(a) + \mu(b) + \mu(c)}, \quad \frac{\mu(c)}{\mu(a) + \mu(b) + \mu(c)}.$$

Определим вероятность того, что хромосома, соответствующая схеме  $s$ , пройдёт в родительский пул. Если в турнир отобраны три хромосомы, соответствующие схеме  $s$ , то с вероятностью равной 1 в родительский пул пройдёт хромосома, соответствующая схеме  $s$ .

Если в турнир отобраны две хромосомы  $a_i, a_j, 1 \leq i, j \leq y$ , соответствующие схеме  $s$ , а третья  $b_k, 1 \leq k \leq M-y$ , не соответствующая схеме  $s$ , то искомая вероятность равна

$$\frac{3y^2(M-y)}{M^3} \cdot \frac{\mu(a_i) + \mu(a_j)}{\mu(a_i) + \mu(a_j) + \mu(b_k)},$$

и средняя вероятность того, что хромосома  $a_i, 1 \leq i \leq y$ , соответствующая схеме  $s$ , пройдёт в родительский пул будет равной  $\frac{3}{M^3} \sum_{\substack{1 \leq i, j \leq y \\ 1 \leq k \leq M-y}} \frac{\mu(a_i) + \mu(a_j)}{\mu(a_i) + \mu(a_j) + \mu(b_k)}.$

Если в турнир отобрана одна хромосома  $a_i, 1 \leq i \leq y$ , соответствующая схеме  $s$ , а две другие  $b_j, b_k, 1 \leq j, k \leq M-y$ , не соответствующие схеме  $s$ , то искомая вероятность равна

$$\frac{3y(M-y)^2}{M^3} \cdot \frac{\mu(a_i)}{\mu(a_i) + \mu(b_j) + \mu(b_k)},$$

и средняя вероятность того, что хромосома  $a_i, 1 \leq i \leq y$ , соответствующая схеме  $s$ , пройдёт в родительский пул  $P$  будет равной  $\frac{3}{M^3} \sum_{\substack{1 \leq i \leq y \\ 1 \leq j, k \leq M-y}} \frac{\mu(a_i)}{\mu(a_i) + \mu(b_j) + \mu(b_k)}.$

Следовательно, математическое ожидание количества хромосом родительского пула соответствующих схеме  $s$  на следующем шаге будет равно

$$E(y(s, t+1)) = y(s, t) \left( \frac{y^3(s, t)}{M^3} + \frac{3}{M^3} \sum_{\substack{1 \leq i, j \leq y(s, t) \\ 1 \leq k \leq M-y(s, t)}} \frac{\mu(a_i) + \mu(a_j)}{\mu(a_i) + \mu(a_j) + \mu(b_k)} + \frac{3}{M^3} \sum_{\substack{1 \leq i \leq y(s, t) \\ 1 \leq j, k \leq M-y(s, t)}} \frac{\mu(a_i)}{\mu(a_i) + \mu(b_j) + \mu(b_k)} \right). \quad (5)$$

Уравнение (5) назовём уравнением репродуктивного роста схемы  $s$  в рассматриваемом случае  $N=3$  турнирно-рулеточной селекции.

Таким образом, при выполнении неравенства



$$\frac{y^3(s,t)}{M^3} + \frac{3}{M^3} \sum_{\substack{1 \leq i, j \leq y(s,t) \\ 1 \leq k \leq M-y(s,t)}} \frac{\mu(a_i) + \mu(a_j)}{\mu(a_i) + \mu(a_j) + \mu(b_k)} + \frac{3}{M^3} \sum_{\substack{1 \leq i \leq y(s,t) \\ 1 \leq j, k \leq M-y(s,t)}} \frac{\mu(a_i)}{\mu(a_i) + \mu(b_j) + \mu(b_k)} > 1$$

схема получает увеличенное число хромосом в следующую генерацию  $G$ .

В частности, из уравнения (5) следует неравенство

$$E(v(s, t+1)) \geq v(s, t) \left( \frac{y^3(s, t)}{M^3} + \frac{6v^2(s, t)(M - y(s, t))}{M^4} \cdot \frac{\bar{\mu}(s, t)}{\bar{F}(t)} + \frac{3v(s, t)(M - y(s, t))^2}{M^4} \cdot \frac{\bar{\mu}(s, t)}{\bar{F}(t)} \right), \quad (6)$$

где  $\bar{\mu}(s, t)$  – средняя приспособленность хромосом, соответствующих схеме  $s$ , в момент  $t$ ,  $\bar{F}(t)$  – средняя приспособленность хромосом всей популяции в момент  $t$ .

Как следует из формулы (6), число строк в популяции, отвечающих схеме  $s$ , растёт, если

$$\frac{\bar{\mu}(s, t)}{\bar{F}(t)} > \frac{M(M^2 + Mv(s, t) + v^2(s, t))}{6v^2(s, t) + 3v(s, t)(M - y(s, t))}. \quad (7)$$

Следовательно, выполнение неравенства (7) является простым достаточным условием роста числа хромосом, соответствующих схеме  $s$ , в следующую генерацию  $G$ .

Оценка математического ожидания  $E(v(s, t+1))$ , получаемая из уравнения (5) репродуктивного роста схемы  $s$  с учётом генетических операторов скрещивания и мутации для турнирно-рулеточной селекции в случае  $N = 3$  имеет вид

$$E(v(s, t+1)) \geq v(s, t) \cdot \left( 1 - p_c \frac{d(s)}{m-1} \right) \cdot (1 - p_m)^{o(s)} \times \\ \times \left( \frac{y^3(s, t)}{M^3} + \frac{3}{M^3} \sum_{\substack{1 \leq i, j \leq y(s, t) \\ 1 \leq k \leq M-y(s, t)}} \frac{\mu(a_i) + \mu(a_j)}{\mu(a_i) + \mu(a_j) + \mu(b_k)} + \frac{3}{M^3} \sum_{\substack{1 \leq i \leq y(s, t) \\ 1 \leq j, k \leq M-y(s, t)}} \frac{\mu(a_i)}{\mu(a_i) + \mu(b_j) + \mu(b_k)} \right), \quad (8)$$

где  $o(s)$  и  $d(s)$  соответственно порядок и определяющая длина схемы  $s$ ,  $p_c, p_m$  – вероятности скрещивания и мутации хромосом из родительского пула  $P$ .

**Случай произвольного  $N < M$ .** В общем случае турнирно-рулеточной селекции аналогичный предыдущим вычислениям подсчёт вероятностей приводит к оценке математического ожидания количества хромосом родительского пула  $P$  на следующем шаге  $E(v(s, t+1))$ :

$$E(v(s, t+1)) \geq v(s, t) \cdot \left( 1 - p_c \frac{d(s)}{m-1} \right) \cdot (1 - p_m)^{o(s)} \times \\ \times \left( \frac{y^N(s, t)}{M^N} + \frac{C_N^{N-1}}{M^N} \sum_{\substack{1 \leq i_k \leq y(s, t), \\ 1 \leq k \leq N-1, \\ 1 \leq j_1 \leq M-y(s, t)}} \frac{\sum_{k=1}^{N-1} \mu(a_{i_k})}{\sum_{k=1}^{N-1} \mu(a_{i_k}) + \mu(b_{j_1})} + \frac{C_N^{N-2}}{M^N} \sum_{\substack{1 \leq i_k \leq y(s, t), \\ 1 \leq k \leq N-2, \\ 1 \leq j_1, j_2 \leq M-y(s, t)}} \frac{\sum_{k=1}^{N-2} \mu(a_{i_k})}{\sum_{k=1}^{N-2} \mu(a_{i_k}) + \mu(b_{j_1}) + \mu(b_{j_2})} + \dots \right. \\ \left. \dots + \frac{C_N^2}{M^N} \sum_{\substack{1 \leq i_1, i_2 \leq y(s, t), \\ 1 \leq j_k \leq M-y(s, t), \\ 1 \leq k \leq N-2}} \frac{\mu(a_{i_1}) + \mu(a_{i_2})}{\mu(a_{i_1}) + \mu(a_{i_2}) + \sum_{k=1}^{N-2} \mu(b_{j_k})} + \frac{C_N^1}{M^N} \sum_{\substack{1 \leq i_1 \leq y(s, t), \\ 1 \leq j_k \leq M-y(s, t), \\ 1 \leq k \leq N-1}} \frac{\mu(a_{i_1})}{\mu(a_{i_1}) + \sum_{k=1}^{N-1} \mu(b_{j_k})} \right), \quad (9)$$



где  $C_n^m = \frac{n!}{m!(n-m)!}$ ,  $o(s)$  и  $d(s)$  соответственно порядок и определяющая длина схемы  $s$ ,  $p_c, p_m$  –

вероятности скрещивания и мутации хромосом из родительского пула  $P$ .

Из оценки (9) соответственно при  $N = 2$  и  $N = 3$  получаются неравенства (4) и (8).

**Работа выполнена при поддержке Российского фонда фундаментальных исследований, грант № 16-29-12911.**

### Список литературы

1. Рутковская Д., Пилиньский М., Рутковский Л. 2013. Нейронные сети, генетические алгоритмы и нечёткие системы. М.: Горячая линия-Телеком.  
Rutkovska D., Pilinsky M., Rutkowski L. 2013. Neural networks, genetic algorithms and fuzzy systems. M: Hotline-Telecom.
2. Кричевский М.Л. 2005. Интеллектуальные методы в менеджменте. СПб.: Питер.  
Krichevsky M.L. 2005. Intelligent methods in management. SPb.: Peter.
3. Гладков Л.А., Курейчик В.В., Курейчик В.М. 2006. Генетические алгоритмы. М.: Физматлит.  
Gladkov L.A., Kureichik V.V., Kureichik V.M. 2006. Genetic algorithms. M.: Fizmatlit. 2006.
4. Holland J. 1975. Adaptation in natural and artificial systems. Ann Arbor, MI: University of Michigan.
5. Petrosov D.A., Lomazov V.A., Dobrunova A.I., Matorin S.I., Lomazova V.I. 2015. Evolutionary synthesis of large discrete systems with dynamic structure. Biosciences Biotechnology Research Asia, V. 12, № 3: 2971–2981.
6. Petrosov D.A., Lomazov V.A., Dobrunova A.I., Matorin S.I., Lomazova V.I. 2015. Large discrete systems evolutionary synthesis procedure. Biosciences Biotechnology Research Asia, V. 12, № 3: 1767–1775.